

車載カメラ映像の俯瞰画像化と意味情報に基づく 道路領域再構成

日大生産工(院) ○柏原 一真 日大生産工 風間 恵介
日大生産工 丸茂 喜高

1. 緒言

自動運転システムの安全性を保障するために、Field Operational Tests(FOT)が世界各国で行われている。自動運転システムの安全性を検証するFOTには、長期間にわたり複数の交通条件や環境条件を観察し、統計的に安全性を保証するものがある。実際に、アルファベット社のウェイモは3,200万kmの実走行試験を行っている¹⁾。しかし、FOTは実際の交通環境で発生する条件を網羅していると保証することができない²⁾。また、安全性は統計的に保証されるため、何十億キロもの走行試験を行う必要がある。既存の調査によると、人間が運転する自動車よりも自動運転システムが運転する自動車の死亡事故に対する安全性が50%改善することを証明するために必要な走行距離は約11億マイル(17.7億km)である³⁾。そのため、統計的な手法のみを用いた安全性の試験はコスト・時間の面で課題となっている⁴⁾。そこで、事故につながる可能性のある状況を分類し、安全性を検証するシナリオベースの安全性評価手法が提案されている⁵⁾。さらに、コンピュータシミュレーションによる自動運転システムの安全性評価も検討されている。これらのことを踏まえるとシミュレーション環境を用いたシナリオベースの安全性評価手法の需要が高いといえる。このような安全性評価シナリオを作成するためには、実際の事故やヒヤリハットシーンから走行した周辺の道路環境を再現し、再現した環境における自車両や他の車両の走行軌跡をデータベース化する取り組みが必要となる⁶⁾。

安全性評価シナリオはドライブレコーダに記録される実際の事故・ヒヤリハットデータを用いて作成することができる。一般に普及しているドライブレコーダに記録されるデータは、車載カメラの映像、Global Navigation Satellite System(GNSS)の測位データ、加速度データである。ヒヤリハットシーンを再現するためには、これらのデータから単に車両の軌跡を推定するだけではなく、車線に対して車両がどのように走行したか推定する必要がある。なぜなら、車両の走行状態が割り込みや追い越し、追突などの状況を判断するためには車線に対する相対的な位置を分析することが必要となるからである。GNSSの測位データと加速度データを用いて位置推定を

行うことができるが、道路状況の再現をすることができない。そのため、本研究では車載カメラに記録される映像を用いた位置推定手法を検討する。単眼カメラを用いた位置推定・マッピングの代表的な手法にVisual SLAMがある。Visual SLAMは映像の特徴を検出し周辺環境をマッピングしながら自己位置推定を行うことができる。しかし、単眼のVisual SLAMには自身の移動量や物体の大きさを定めることができないという課題がある⁷⁾。また、Visual SLAMにおいて作成される地図は計算量削減のために特徴の点群が用いられることが多い。そのため交通状況の可視化という点では適さない。既存の研究では航空衛星画像と前方画像に対して画像位置合わせを用いて車両の位置推定を行う⁸⁾。しかし、画像中の道路領域の遮蔽率が大きい場合位置推定が困難である。そこで、位置推定に用いる計算の対象となる画素を道路領域に限定することで、位置推定精度が向上すると考えた。この道路領域の判別にはSemantic Segmentationを用いることができる。Semantic Segmentationは画像処理技術の1つで、画像の各画素を車両や路面などの意味をもつ画素値に分類する技術である。

本研究では、Semantic Segmentationを用いて得た車両前方画像に含まれる道路領域の情報を使用して道路領域と自車両の走行位置の可視化を目的とした。ヒヤリハットシーンの解析ではドライブレコーダに記録されるデータを解析対象とするため、本研究ではリアルタイムな処理は必要でない。また、ドライブレコーダでは走行位置の真値を得られないため、真値の存在するシミュレータを用いて解析対象とするデータを作成し、提案手法の妥当性を検討した。

2. 提案手法

本研究で用いる自車両の位置推定手法と道路領域の再現方法について記述する。本研究で対象とするデータは自動運転シミュレータCARLA⁹⁾を用いて作成した。車載カメラに記録される前方画像を俯瞰(Bird's Eye View; BEV)画像に変換する。BEV変換には画像サイズとカメラの取付位置とピッチ角、焦点距離、光学的中心の値を用いた。BEV変換は自車両の重心がBEV画像の下辺の中点となるように行った。作成したBEV画像は自車

両の重心を基準に前方12.8m, 左右6.4mの範囲の情報を持つ。作成したBEV画像のサイズは256x256pixelであり, 解像度は0.05m/pixelである。画像位置合わせを用いることで連続する2フレームのBEV画像が示す領域がどれだけ移動したかを算出することができる^{10,11)}。走行路面が平面であると仮定すると, BEV画像のフレーム間の移動量を自車両の移動量とみなすことができる。しかし, 画像位置合わせはBEV画像中に動的物体や高さ情報を含む画素が多く存在すると誤った位置合わせが多くなる。そこで, Semantic Segmentationを用いて道路領域の情報のみをもつマスク画像を作成し, 画像位置合わせの計算領域を指定した画像位置合わせを用いて推定したフレーム間の移動量を用いてBEV画像を貼り合わせ道路領域のみの情報を持つ地図を作成した。

2.1. 計算領域を指定するBEV画像の作成

道路領域の情報のみをもつマスク画像を作成し, 画像位置合わせの計算領域を指定する。図1は, 計算領域の画像を作成するための構成要素を示す。

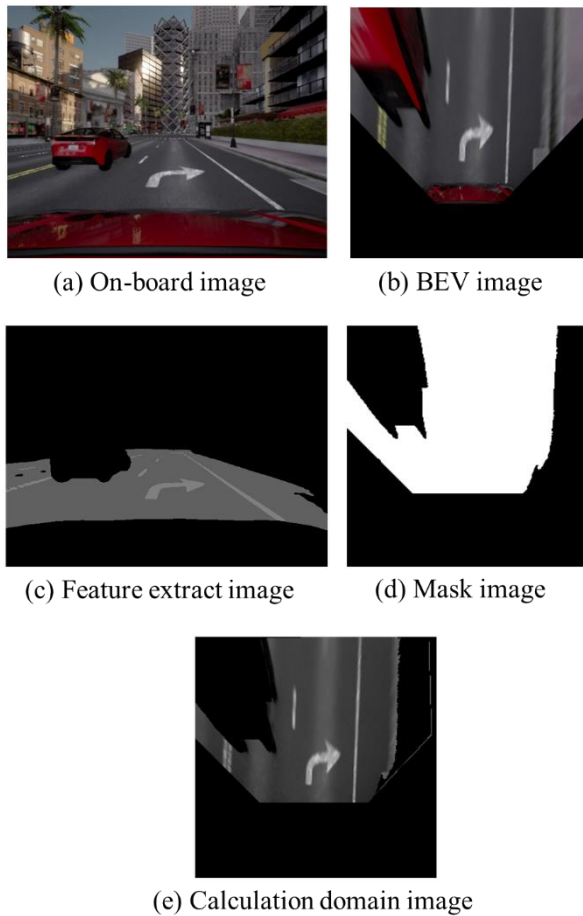


Fig. 1 Components of creating images of the calculation domain

図1(a)に示す前方画像に対してBEV変換を行い図1(b)に示す前方画像のBEV画像を作成する。次に, 図1(a)示した前方画像に対してSemantic Segmentationを用いて, 図1(c)に示す3種類の画素値の画像を作成した。それぞれの画素値は路面標示を128, 道路領域を96, それ以外の領域を0とした。Semantic SegmentationのモデルにはHR-Net+OCR¹²⁾を用いた。使用した重みは, CityScapesデータセット¹³⁾を用いた学習済みの重みに対してA2D2のデータセット¹⁴⁾を用いて転移学習を行うことで車線検出を可能にしたものを用いた。道路標示と道路領域の特徴を抽出した画像に対して, 射影変換を行いBEV画像作成した。図1(d)に示すマスク画像は, このBEV画像の路面標示・道路領域の画素値を255それ以外の画素値を0とし, 下部80pixelは自車両のボンネット部分であるため画素値を0として作成した。図1(b)のBEV画像に図1(d)のマスク画像を用いてマスク処理することで図1(e)に示す路面標示・道路領域のみの情報をもつBEV画像を作成した。画像位置合わせの計算領域とした画素は図1(e)に示したBEV画像の画素値が0でない領域である。

2.2. 画像位置合わせ

計算領域を指定したフレーム番号 n のBEV画像を $M(x)$, 計算領域を指定したフレーム番号 $(n+1)$ のBEV画像を $F(x)$ とする。 $M(x)$ と $F(x)$ に対して画像位置合わせを用いることでフレーム間の自車両の推定移動量を求める。フレーム間の自車両の推定回転移動量を T_θ , 推定平行移動量を T_x, T_y とする。 $F(x)$ を固定し, $M(x)$ を後述する探査範囲内で線形変換して移動させる。線形変換に用いる画像の回転・平行移動パラメータを $\mathbf{p} = (T_x, T_y, T_\theta)^T$, 画像のピクセル座標を示すベクトルを $\mathbf{x} = (x, y)^T$ とする。線形変換後のピクセル座標 $\mathbf{x}' = (x', y')^T$ は式(1)によって求めることができる。式(1)を用いることで画像 $M(\mathbf{x}, \mathbf{p})$ を得る。これは線形変換後のフレーム番号 n のBEV画像である。

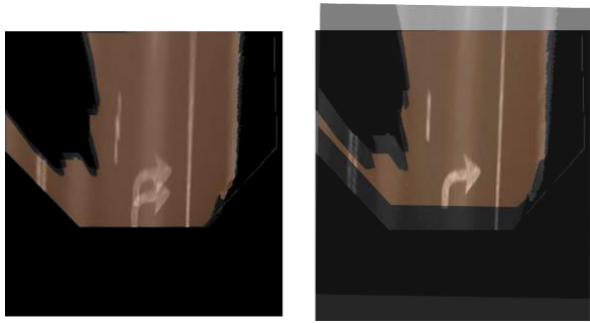
$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{pmatrix} \cos T_\theta & -\sin T_\theta & T_x \\ \sin T_\theta & \cos T_\theta & T_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \quad (1)$$

画像位置合わせ前後の2枚の画像の位置関係を図2に示す。図2(a)に示した画像は位置合わせ前であるため, 画像間における路面標示のずれがわかる。画像位置合わせは, 各画素値の差を最小化するように画像を移動させるため, 図2(b)に示した画像は道路標示のずれが減少していることが確認できる。図2のオレンジ色の領域は $F(x)$ と $M(x, \mathbf{p})$ の計算対象となる領域が重なる部分を示す。この重なる部分に対し, 画素値のRoot Mean Squared Error(RMSE)を計算し, RMSEが最も小さくなる \mathbf{p} をフレーム間の推定移動量とした。共通領域のピ

クセル数を N 、画像間で対応する任意のピクセル座標の画素値をそれぞれ $(F(\mathbf{x}_i), M(\mathbf{x}, \mathbf{p})_i)$ ($i = 0, \dots, N$)とする。画像位置合わせの誤差関数 E を最小化する \mathbf{p} は式(2)によって求められる。

$$E(\mathbf{p}) = \frac{1}{N} \sum_{i=0}^N (M(\mathbf{x}, \mathbf{p})_i - F(\mathbf{x}_i))^2 \quad (2)$$

\mathbf{p} を定めるために探査する範囲は車両運動が非ホロノミック拘束に制限されると仮定して定めた。フレーム間のヨー角の最大変化量を 24deg/s と仮定することで T_θ の探査範囲は $-1.6\sim 1.6\text{deg}$ とした。平行移動量の探査範囲は最高時速を 30.0m/s (108km/h)と仮定した。 T_x の探査範囲は $0\sim 40\text{pixel}$ 、 T_y の探査範囲は $-2\sim 2\text{pixel}$ の範囲とした。また、探査分解能は回転移動量が 0.2deg 刻み、平行移動量 1pixel (0.05m)刻みとした。



(a) Before image registration (b) After image registration

Fig. 2 Image position before and after image registration

2.3. 自車両の位置推定及び地図作成

推定したフレーム間の移動量を用いて自車両の位置を算出する。車両の初期姿勢および初期位置は既知として与えた。地図座標系での n フレーム目の自車両の位置座標とヨー角をそれぞれ (X_n, Y_n) と φ_n とする。 $(n+1)$ フレーム目の自車両の位置と姿勢は式(3)によって求めることができる。

$$\begin{pmatrix} X_{n+1} \\ Y_{n+1} \\ \varphi_{n+1} \end{pmatrix} = \begin{pmatrix} X_n \\ Y_n \\ \varphi_n \end{pmatrix} + \begin{pmatrix} T_x \cos(\varphi_n + T_\theta) - T_y \sin(\varphi_n + T_\theta) \\ T_x \sin(\varphi_n + T_\theta) + T_y \cos(\varphi_n + T_\theta) \\ T_\theta \end{pmatrix} \quad (3)$$

算出した n フレーム目の位置座標 (X_n, Y_n) に n フレーム目の道路領域のみの情報をもつBEV画像を張り合わせることで地図画像を作成する。このとき、画素値が0の領域は情報を持たないため、0以外の画素値の場合のみ画素値を上書きする。このことにより、路面を遮蔽する物体がある場合

においても、いずれかのBEV画像で道路領域の情報が得られれば地図画像に描画することができる。

3. 検証結果

提案手法の妥当性を検討するため、自動運転シミュレータCARLAを用いて曲線区間を含む走行データを作成した。車両前方映像を取得する車載カメラの内部パラメータは既知とした。車載カメラは車両重心を基準に前方 1m 、道路路面より上方 1.3m の位置に取り付けた。車両前方映像の解像度は $640 \times 480\text{pixels}$ 、視野角 90deg 、フレームレート 15fps とした。作成したデータの代表的な前方画像を取得した時刻とともに図3に示す。作成したデータは自車両が静止した状態で開始し、左前方の他車両と並走する。シミュレーション開始から 1.0s 経過時点で加速を始める。 8.9s 経過時点で最高時速 16.4m/s (59.3km/h)に到達し、その後減速動作を行う。 13.0s 経過時点で横断歩道の前で停止する。



Fig. 3 Simulation data containing curved section

図4にシミュレータを用いて作成したデータから再現した道路領域を可視化し、オレンジ色の点を用いて自車両の位置推定結果を示す。地図画像も路面を遮蔽する物体がある場合でも、いずれかのBEV画像で道路領域の情報が得られれば地図画像に描画することができた。そうでない場合は黒色の失われた領域が発生した。

4. まとめ

本研究では、車載カメラから得られる前方画像を用いて道路領域と自車両の走行位置の可視化を目的とした。位置推定に用いる計算の対象となる画素を道路領域に限定するためにSemantic Segmentationを用いた。また、求めた移動量をもとに、走行した道路領域と自車両の走行位置の再現を行った。提案手法により道路領域のみの情報を持つ地図画像を作成することができた。また、推定した自車両の位置を作成した地図画像上に描画し可視化することができた。今後は、各フレーム間における位置推定誤差の累積軽減を図り、相対距離推定することで他車両の軌跡を推定し、より現実的な交通状況の可視化を行う。

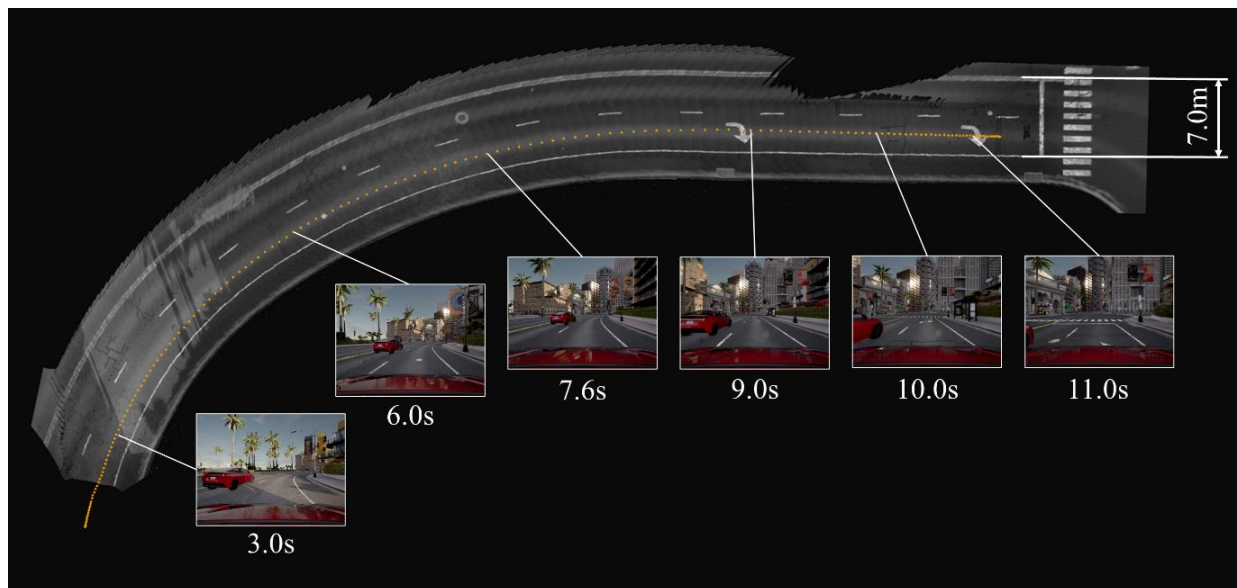


Fig. 4 Road region visualization and position estimation results

参考文献

- 1) P. Sun, et al., “Scalability in perception for autonomous driving: Waymo open dataset”, Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, (2020) pp.2446-2454.
- 2) 一般社団法人 日本自動車工業会 自動運転部会 AD 安全性評価分科会, 自動運転の安全性評価フレームワーク Ver 3.0, (2022).
- 3) N. Kalra and S. M. Paddock, “Driving to safety: How many miles of driving would it take to demonstrate autonomous vehicle reliability?”, Transportation Research Part A: Policy and Practice, Vol. 94, (2016) pp.182-193.
- 4) W. Wachenfeld and H. Winner, “The Release of Autonomous Vehicles”, Autonomous Driving: Technical, Legal and Social Aspects, (2016) pp.425-449.
- 5) S. Ulbrich, et al., “Defining and substantiating the terms scene, situation, and scenario for automated driving”, In 2015 IEEE 18th international conference on intelligent transportation systems, (2015) pp.982-988.
- 6) Y. Akagi, “Ontology based collection and analysis of traffic event data for developing intelligent vehicles”, In 2017 IEEE 6th Global Conference on Consumer Electronics (GCCE), (2017) pp.1-5.
- 7) D. Chekhlov, et al., “Robust real-time visual slam using scale prediction and exemplar based feature description”, In 2007 IEEE conference on computer vision and pattern recognition, (2007) pp.1-7.
- 8) M. Noda, et al., “Vehicle ego-localization by matching in-vehicle camera images to an aerial image”, In Computer Vision-ACCV 2010 Workshops: ACCV 2010 International Workshops, Revised Selected Papers, Part II 10, (2011) pp.163-173.
- 9) A. Dosovitskiy, et al., “CARLA: An open urban driving simulator”, Proceedings of the 1st Annual Conference on Robot Learning, (2017) pp.1-16.
- 10) D. B. Lucas and T. Kanade, “An Iterative Image Registration Technique with an Application to Stereo Vision”, Proceedings of 7th Joint Conference on Artificial Intelligence (IJCAI), (1981) pp.674-679.
- 11) S. Baker and I. Matthews, “Lucas-Kanade 20 years On: A Unifying Framework”, International Journal of Computer Vision, Vol. 56, No. 3, (2004) pp.221-255.
- 12) Y. Yuan, et al., “Object-contextual representations for semantic segmentation”, In Computer Vision-ECCV 2020: 16th European Conference, Proceedings, Part VI 16, Springer International Publishing, (2020) pp.173-190.
- 13) M. Cordts, et al., “The Cityscapes Dataset for Semantic Urban Scene Understanding”, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), (2016) pp.3213-3223.
- 14) J. Geyer, et al., “A2d2: Audi autonomous driving dataset”, arXiv preprint arXiv:2004.06320, (2020).